

A GENERAL REVIEW OF DATA MINING PATTERNS

SUSMITA B. DAS¹ & MAYURI V. CHAUDHARI²

¹Department of Computer Engineering, B. R. Harne College of Engineering, Karav, Vangani, Ambernath, Thane,
Maharashtra, India

²Department of Electronics and Telecommunication Engineering, B. R. Harne College of Engineering, Karav, Vangani,
Ambernath, Thane, Maharashtra, India

ABSTRACT

"Mining" is the extraction of valuable materials from the core of the earth which are of great economic interest or importance. Traditionally, mining has been used at excavation sites for extraction of minerals like gold and copper. Data mining comprises of unearthing useful patterns from a data warehouse which is the source of integrated data. Data mining can also be used as a BI (Business Intelligence) tool to predict or derive useful patterns by the analysis of current and historical data. In a broader scope it is an inter-disciplinary subfield of computer science and makes use of the computational processes like Machine Learning, statistics, Artificial Intelligence and database systems to discover patterns and make it available in a human readable format for prediction of future trends.

KEYWORDS: Data Mining, Data Analysis

INTRODUCTION

Data mining involves automatic or semi-automatic analysis of data records (by using cluster analysis) and dependencies (association rule mining). For example, data mining step can be used to identify multiple groups in data records which can be then used to predict more accurate results for decision support systems (DSS). Data analysis/mining of statistical information collected by the government like population statistics can help to establish demographic information of a particular area. This information can then be used for deciding the type of E-services to be provided under the E-governance vision in that area. Perhaps, a simple example in this scenario could be provision of E-services based on anyone of the population characteristics example, age groups. Based on this information the government can organize medical health checkup facilities or improvise on the medical facilities if the general population of that area falls into the senior citizen group. If the population characteristics data on age groups, indicates a good heterogeneity of various groups like children in the age group of 3-10 years, young population of 15-30 year olds as well as senior citizens then we can predict that the facilitation of infrastructural services like new schools, hospitals, free vaccination drives for children, infrastructure like new colleges, government shops for issuance of transport certificates, ration cards, passports, payment of bills, for the young population all under one roof and free medical aid for senior citizens would be beneficial for the citizens.

ANALYSIS OF DIFFERENT PATTERNS IN DATA MINING

A useful pattern that can be obtained after data analysis of purchases by customers could be something like the Bread-Butter theory its analysis could be as follows, considering a grocery store which has list of food products like bread, butter, biscuits, cheese, milk and eggs. Records of purchases by customers are maintained. An analysis of a few shopping

records registers the point that most customers who brought bread as a shopping item in the first place have brought cheese, eggs, butter or milk as a secondary shopping item, thus bread is associated with any of the products or customers buying bread are showing a certain probability of buying any of the items specified. The probabilities of the secondary items brought along with bread could be established by evaluating a record of 100 shopping lists, we verify that the probability associated with butter is 0.5, eggs is 0.2, milk is 0.2 and cheese is 0.1. Out of these statistics the associations of bread with butter are the strongest verifying that most customers of the store prefer buying bread and butter together so these two products could also be sold together and placing the two items together will result in the better sale of the two products and hence imply better revenues. Also the patterns obtained like bread-eggs and bread-milkal though of lower probabilities are relevant associations and can be grouped together so that the products could be sold together as a unit and certain discounts be placed on the combinations so that more products fly off the shelves. Also, these associations obtained could be applied on other shopping stores of the same organization with the intent of maximizing profits. Demographic data obtained from population statistics is also a crucial factor in decision support systems in various regards. If the population characteristics like age, income group, standards of living of a particular area can be obtained then the decision regarding the types of products that can be sold in that particular area can also be established. A comparative analysis of a shopping store in the wealthy suburbs of Mumbai and a similar store in a mid-income zone of Mumbai in terms of sales or revenues generated can help taking decisions as to opening of a new business venture in either of the two places. Thus, data mining can be used for decision support after performing a cost-to-benefit analysis.

EVENT PROCESSING AND DATA MINING FOR SMART CITIES

In a city various events might take place for example a musical event. If data sources related to the city which include traffic characteristics, temperature, cultural events are gathered over time then certain association rules can be formulated which can help in the creation of smart cities which will be able to deal with these events in a more appropriate manner. A typical scenario would be organization of a music event on a weekend which can draw up to thousands of viewers. Would data mining predictions be useful in this case? Yes, previous data related to traffic characteristics can help in the usage of alternative routes to disperse traffic on weekends when events are organized thus bringing about efficient utilization of resources and time (example, highways to be used instead of general routes) and also eradicating unnecessary traffic snarls and inconvenience.

CONCLUSIONS

Although the scope and applications of data mining is still in its nascent phases the enormous potential of the techniques and tools for data mining can be used in a wide variety of industries- retail, manufacturing, transportation to name a few that is any organization which has data be it of current or archived nature and wishes to analyze this data can use it for obtaining strategic patterns, facts, relationships, trends, exceptions and anomalies that might otherwise go unnoticed.

REFERENCES

1. Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S. K. Ratnam, "A Study of Data Mining Tools in Knowledge Discovery Process" *Volume-2*, July 2012.
2. J. Han and M. Kamber, "Data Mining: Concepts and Techniques", 2000.
3. Alexandra Moraru, Dunja Mladenić "Complex Event Processing and Data Mining for Smart Cities".